On Length Independent Security Bounds for the PMAC Family

Bishwajit Chakraborty[1], Soumya Chattopadhyay[1], Ashwin Jha[2] and Mridul Nandi[1]

[1] Indian Statistical Institute, Kolkata, India
{bishu.math.ynwa,s.c.2357,mridul.nandi}@gmail.com
[2] CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
ashwin.jha@cispa.de

Abstract. At FSE 2017, Gaži et al. demonstrated a pseudorandom function (PRF) distinguisher (Gaži et al., ToSC 2016(2)) on PMAC with $\Omega(\ell q^2/2^n)$ advantage, where q, ℓ, and n, denote the number of queries, maximum permissible query length (in terms of n-bit blocks), and block size of the underlying block cipher. This, in combination with the upper bounds of $O(\ell q^2/2^n)$ (Minematsu and Matsushima, FSE 2007) and $O(q\sigma/2^n)$ (Nandi and Mandal, J. Mathematical Cryptology 2008(2)), resolved the long-standing problem of exact security of PMAC. Gaži et al. also showed that the dependency on ℓ can be dropped (i.e. $O(q^2/2^n)$ bound up to $\ell \leq 2^{n/2}$) for a simplified version of PMAC, called sPMAC, by replacing the Gray code-based masking in PMAC with any 4-wise independent universal hash-based masking. Recently, Naito proposed another variant of PMAC with two powering-up maskings (Naito, ToSC 2019(2)) that achieves ℓ-free bound of $O(q^2/2^n)$, provided $\ell \leq 2^{n/2}$. In this work, we first identify a flaw in the analysis of Naito's PMAC variant that invalidates the security proof. Apparently, the flaw is not easy to fix under the existing proof setup. We then formulate an equivalent problem which must be solved in order to achieve ℓ-free security bounds for this variant. Second, we show that sPMAC achieves $O(q^2/2^n)$ bound for a weaker notion of universality as compared to the earlier condition of 4-wise independence. Third, we analyze the security of PMAC1 (a popular variant of PMAC) with a simple modification in the linear combination of block cipher outputs. We show that this simple modification of PMAC1 has tight security $O(q^2/2^n)$ provided $\ell \leq 2^{n/4}$. Even if $\ell > 2^{n/4}$, we still achieve same tight bound as long as total number of blocks in all queries is less than $2^{2n/3}$.

Keywords: PMAC · PMAC1 · PMAC_Plus · PRF · universal hash · tight security

1 Introduction

Message Authentication Codes or MACs are symmetric-key primitives that ensure data integrity and authenticity. PMAC, by Black and Rogaway [BR02], is an example of parallelizable block cipher-based MAC. A slightly simplified version[1] of PMAC based on an n-bit block cipher E_K is defined as follows:

$$\mathsf{PMAC}_K(m) := E_K\left(E_K(m_1 \oplus \gamma_1 \cdot \Delta) \oplus \cdots \oplus E_K(m_{\ell-1} \oplus \gamma_{\ell-1} \cdot \Delta) \oplus m_\ell\right),$$

where $(m_1, \ldots, m_\ell)$ is n-bit (also referred as block) parsing of the input message m, $(\gamma_1, \cdots, \gamma_{\ell-1})$ is the gray code sequence [Gra53, Rog04] and $\Delta = E_K(0^n)$ is the masking

[1] Ignoring the padding rule.

key. PMAC and its close variant PMAC1 [Rog04] have the ability to significantly outperform sequential block cipher based-MACs, like the CBC-MAC family [EMST76, BKR94, BdBB+95, BR00], by virtue of their parallelizable nature.

Existing Analysis of PMAC: In the following discussion, q, ℓ and σ, respectively, denote the number of queries, maximum permissible message length in blocks, and total number of blocks in all queries, i.e, $\sigma \leq \ell q$. It is a well-known observation [GGM84, BGM04] that a good PRF is necessarily a good deterministic[2] MAC. Consequently, most of the research on the security of PMAC have explored its pseudorandomness properties. The first result along this line came in the introductory paper by Black and Rogaway [BR02] who showed an upper bound of $O(\sigma^2/2^n)$ on the PRF advantage. A different bound of the form $O(\ell q^2/2^n)$ was shown by Minematsu and Matsushima [MM07]. This bound is better than the original bound whenever message lengths do not vary much from ℓ. However, this bound can be worse when very few messages are of length ℓ and rest of the messages are of length much smaller than ℓ. Nandi and Mandal [NM08] showed an improved bound about $O(q\sigma/2^n)$. This is indeed an improved bound for all choices of parameters.

Luykx et al. [LPSY16] studied the problem from lower bound perspective. Specifically, they constructed a pair of messages such that the PMAC outputs corresponding to the two messages collide with probability roughly $\ell/2^n$, leading to a distinguishing attack with advantage $\ell/2^n$ for $q = 2$ queries. However, they did not show how this can be extended to get collision probability about $\ell q^2/2^n$ for $q \geq 2$ messages. Later Gaži et al. [GPR16] constructed an adversary which makes q queries, each of length exactly ℓ blocks, so that the collision probability of PMAC outputs is about $\ell q^2/2^n$. Thus, the bounds $\ell q^2/2^n$ and $q\sigma/2^n$ are essentially tight. However, it is worth noting that the attack does not work for PMAC1 [Rog04] where the Gray code sequence is replaced with the sequence $\alpha, \alpha^2, \alpha^3, \ldots$ for some fixed primitive element α of the Galois field GF(2^n). So, the exact security of PMAC1 is still an open problem.

PRFs with Length Independent Security: In applications where we process large messages or where most of the messages are of lengths much smaller than ℓ, a bound of the form $O(q^2/2^n)$ (length-independent) is much desired, as compared to say a bound of $O(\ell q^2/2^n)$. For instance, AES128 [NIS01] based PMAC needs rekeying after roughly 2^{22} messages when message length can be as large as 2^{56} bytes and more than 2^{-32} advantage is not tolerated. On the other hand, any construction with $q^2/2^n$ or similar bound can be safely used without rekeying for up to 2^{48} messages in a similar setup. As a result, this line of research has seen a lot of interest over the years.

EMAC [BKR94, BdBB+95], ECBC and FCBC [BR00] are shown to have $O(q^2/2^n)$ PRF advantage provided $\ell \leq 2^{n/4}$ [JN16a, JN16b]. However, these constructions are sequential in nature. Luykx et al. [LPTY16] proposed a parallel construction, called LightMAC, that achieves ℓ-free security. However, inspired by Bernstein's protected counter sums [Ber99], LightMAC uses a counter-based encoding which limits the efficiency. For example, to allow a message length of $2^{n/2}$ blocks, LightMAC requires two calls of block ciphers to process one block of message, i.e., it is a rate[3] 1/2 construction. Dutta et al. [DJN17] proposed some optimal strategies to encode counter and message in input blocks. Although this increases the rate for smaller messages, still the rate is low as compared to PMAC or PMAC1.

With respect to PMAC-like designs, Gaži et al. [GPR16] proved $O(q^2/2^n)$ bound for a simplified variant of PMAC, called sPMAC, albeit with comparatively expansive masking methods. For example, the masking function should be a 4-wise independent function. Most efficient algebraic instantiations of such a function require at least four keys and several field multiplications. Very recently, Naito [Nai19] proposed a variant of PMAC1,

[2] This observation is not true, in general, for nonce-based or probabilistic MACs.

[3] Roughly speaking, rate is the ratio of message length in blocks to the number of block cipher calls required to process the message.

which uses two powering-up maskings (instead of one used in PMAC1). He showed $O(q^2/2^n)$ advantage provided $\ell \leq 2^{n/2}$.

The constructions following Double-block Hash-then-Sum paradigm [DDNP18], including PMAC_Plus [Yas11] and LightMAC_Plus [Nai17], achieve beyond the birthday bound (BBB) security [KLL20] and hence can achieve ℓ-free bound for a wide range of ℓ. However, these constructions require almost twice the memory (due to the BBB security requirement) used in other PMAC variants. So, in this paper we only focus on PMAC-like designs that follow the Hash-then-PRP paradigm [Sho04].

1.1 Our Contributions

Table 1: A comparative summary of several PMAC variants. Here q denotes the number of queries, ℓ denotes the upper bound on query-length, and σ denotes the upper bound on total number of blocks present in all queries.

Mode	Security bound	Length restriction	Number of masking keys
PMAC [BR02]	$q^2\ell/2^n$	-	1
PMAC1 [Rog04]	$q^2\ell/2^n$	-	1
NPMAC[1] [Nai19]	$q^2/2^n$	$\ell < 2^{n/2}$	2
PMAC3 [Nai20]	$q^2/2^n$	$\ell < 2^{n/2}$	3
PMAC2 [Section 7]	$q/2^{n/2}$	$\ell \leq 2^{n/4}$	1
	$\sigma^{1.5}/2^n$	$2^{n/4} < \ell \leq 2^{n-2}$	1

[1] The security analysis of this construction is shown to be incorrect in this paper.

Our contributions are threefold:

1. Revisiting Naito's variant of PMAC1: As of now, Naito's PMAC1 variant [Nai19], sometimes also referred as NPMAC in this paper, is the only known rate-1 PMAC-like construction that achieves ℓ-free security bound (for $\ell < 2^{n/2}$). We show that *the security analysis of this construction is incorrect* (see Section 4). Further, we state an equivalent problem which must be solved to prove the ℓ-free security of this construction. However, we are not able to solve that equivalent problem. So the exact security of Naito's variant is still an open problem. Naito subsequently updated the construction [Nai20] in light of our observations. This updated variant achieves ℓ-free security for $\ell < 2^{n/2}$ (see Section 5).

2. Relaxing the Security Precondition for sPMAC: In [GPR16], sPMAC is shown to have ℓ-free security bound up to $\ell < 2^{n/2}$ when the underlying masking function is 4-wise independent hash. We *relax the* 4*-wise independence condition to* 2*-wise almost XOR universality* (see Section 5).

3. PMAC2 – A simple variant of PMAC1: As we still lack of an ℓ-free secure PMAC variant with efficient masking function, our next part is aimed to solve this problem. We propose a simple variant of PMAC1, called PMAC2, and we show almost tight security $O(q/2^{n/2})$ (see Table 1). More precisely, we prove the following theorem (in Section 7).

   Security Analysis of PMAC2: Let ℓ denote the number of blocks present in the longest query and σ denotes the total number of blocks present in q queries altogether.

Then,

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{PMAC2}}(q,\ell,\sigma) \leq \frac{2q^2+\sigma}{2^n} + \mu$$

where $\mu \leq \frac{q}{2^{n/2}}$ if $\ell \leq 2^{n/4}$ and $\mu \leq \frac{\sigma^{1.5}}{2^n}$ if $2^{n/4} < \ell \leq 2^{n-2}$.

2 Preliminaries

Basic Notations: For any positive integer n, we write $[n] := \{1,\ldots,n\}$. We write x^q to denote a q-tuple $(x_1,\ldots,x_q)$. We write $X \leftarrow\!\!\$\ \mathcal{X}$ to represent that X is a uniform random variable taking values from a finite nonempty set $\mathcal{X}$.

Throughout, $\boldsymbol{\rho}_{\mathcal{D}} \leftarrow\!\!\$\ \mathsf{Func}_{\mathcal{D}}$ denotes a random function, and $\boldsymbol{\pi} \leftarrow\!\!\$\ \mathsf{Perm}$ denotes a random permutation. We simply write the random function as $\boldsymbol{\rho}$, when $\mathcal{D}$ is understood from the context.

Notations on Blocks: Throughout the paper n denotes the security parameter as well as the bit size of the underlying permutation. We call the set $\mathfrak{B} := \{0,1\}^n$ block set and elements of the set *blocks*. We define $\mathfrak{B}^+ = \cup_{i\geq 1}\mathfrak{B}^i$. For any binary string $m \in \{0,1\}^*$, we denote the number of bits of m as $|m|$ and we write $\|m\| := \lceil |m|/n \rceil$.[4] We use "$\|$" to denote concatenation operations on bit strings. For a message $m \in \{0,1\}^{nl}$, we write $m = m[1]\|\cdots\|m[l]$ with $m[i] \in \{0,1\}^n$ for all $i \in [l]$.

Notations on Block Functions and Permutations: We call a function block function if the range of the function is the block set. The set of all block functions defined over a set $\mathcal{D}$ is denoted as $\mathsf{Func}_{\mathcal{D}}$. The set of all permutations over the block set (also called block permutation) is denoted as Perm.

A keyed block function F with key space $\mathcal{K}$ and domain $\mathcal{D}$ is a block function over $\mathcal{K}\times\mathcal{D}$. We also view it as an indexed family of functions, where $\mathcal{K}$ is the index set, i.e., for each $K \in \mathcal{K}$, we associate a function $F_K(\cdot) := F(K,\cdot)$.

Multiset: Informally, a multiset $\mathcal{X}$ is a variant of set in which we allow elements to repeat. One can equivalently define a multiset $\mathcal{X}$ by a set $\{(x,m) : x \in \mathcal{X}, x \text{ appears } m \text{ times in } \mathcal{X}\}$. We write $\mathcal{X}^o$ to denote the set of all elements x which appears odd times in $\mathcal{X}$. Note that, $\mathcal{X}^o$ by definition is a set which can be empty. We say $\mathcal{X}$ is **evenly repeated** if $\mathcal{X}^o = \emptyset$.

Example 1. Let $\mathcal{X} := \{a,b,a,b,b,c\}$ be a multiset. We represent it by the following set $\{(a,2),(b,3),(c,1)\}$. Note that $\mathcal{X}^o = \{b,c\}$. Similarly, for a multiset $\mathcal{Y} := \{a,b,a,b,b,b,c,c\}$, $\mathcal{Y}^o = \emptyset$ and hence $\mathcal{Y}$ is evenly repeated.

Given a block function π, we use shorthand notation $\pi^{\oplus}(\mathcal{X}) := \bigoplus_{x\in\mathcal{X}} \pi(x)$. With this notation, it is easy to see that (the empty sum represents 0^n)

$$\pi^{\oplus}(\mathcal{X}) = \pi^{\oplus}(\mathcal{X}^o) \text{ for every multiset } \mathcal{X}, \tag{1}$$

and hence $\pi^{\oplus}(\mathcal{X}) = 0^n$ whenever $\mathcal{X}$ is evenly repeated multiset.

Binary Field: In this paper, we view the block set $\mathfrak{B}$ as the Galois field $\mathrm{GF}(2^n)$. We fix a primitive polynomial $p(x) := p_0 \oplus p_1 x \oplus \cdots \oplus p_n x^n$ where $p_i \in \{0,1\}$. Note that $p_0 = p_n = 1$ (as it is a primitive polynomial). The field multiplication "$\cdot$" between two field elements is defined through the primitive polynomial. We abuse the notation 2 to denote a primitive element of the underlying field $\mathrm{GF}(2^n)$.

[4] When m is a set we also write $|m|$ to denote the size of the set m. So the notation $|m|$ should be clear from the context.

2.1 Hash Functions

In the following, let H be a keyed block function with keyspace $\mathcal{K}$ and domain $\mathcal{D}$.

Collision Probability: For distinct $m, m' \in \mathcal{D}$, we define collision probability as

$$\mathsf{coll}_H(m, m') := \mathsf{Pr}(H(K, m) = H(K, m') : K \leftarrow\!\!\$\, \mathcal{K}).$$

When $\mathcal{D} \subseteq \{0,1\}^*$, the collision probability can depend on the size of the inputs. We write

$$\mathsf{coll}_H(\ell) = \max_{\substack{m \neq m' \\ |m|,|m'| \leq \ell}} \mathsf{coll}_H(m, m').$$

We generalize the above definition for more than two inputs. For q distinct inputs $m_1, \ldots, m_q \in \mathcal{D}$, we write

$$\begin{aligned} \mathsf{coll}_H(m^q) &:= \mathsf{Pr}(\exists i < j, H(K, m_i) = H(K, m_j) : K \leftarrow\!\!\$\, \mathcal{K}), \text{ and} \\ \mathsf{coll}_H(q, \ell, \sigma) &:= \max_{\substack{m^q : |m_i| \leq \ell \\ \sum_{i=1}^{q} |m_i| \leq \sigma}} \mathsf{coll}_H(m^q). \end{aligned}$$

By using the union bound, $\mathsf{coll}_H(q, \ell, \sigma) \leq \binom{q}{2} \mathsf{coll}_H(\ell)$.

Definition 1 (Universal hash function)**.** The keyed block function H is called an ϵ-universal hash if for all distinct $m, m' \in \mathcal{D}$, $\mathsf{coll}_H(m, m') \leq \epsilon$.

Definition 2 (XOR universal hash function)**.** The keyed block function H is called an ϵ-almost XOR universal hash if for all distinct $m, m' \in \mathcal{D}$ and $\delta \in \mathfrak{B}$,

$$\mathsf{Pr}(H(K, m) \oplus H(K, m') = \delta : K \leftarrow\!\!\$\, \mathcal{K}) \leq \epsilon.$$

Definition 3 (k-wise independent hash function)**.** The keyed block function H is called a k-wise independent if for all distinct $m_1, \ldots, m_k \in \mathcal{D}$ and for all $y_1, \ldots, y_k \in \mathfrak{B}$,

$$\mathsf{Pr}(H(K, m_1) = y_1, \ldots, H(K, m_k) = y_k : K \leftarrow\!\!\$\, \mathcal{K}) = \frac{1}{2^{kn}}.$$

The following observations are easy to establish.

1. A random function is k-wise independent for any k.
2. A 2-wise independent hash function is 2^{-n}-AXU.

2.2 Pseudorandom Functions and the Hash-then-RP Paradigm

Definition 4 (Pseudorandom function)**.** Let F be a keyed block function over a finite set $\mathcal{D}$ with a finite key space $\mathcal{K}$. The *PRF-advantage* of any oracle adversary $\mathcal{A}$ against F is defined as

$$\mathbf{Adv}^{\mathrm{prf}}_F(\mathcal{A}) := \left| \mathsf{Pr}(\mathcal{A}^{F_K} = 1 : K \leftarrow\!\!\$\, \mathcal{K}) - \mathsf{Pr}(\mathcal{A}^{\boldsymbol{\rho}_{\mathcal{D}}} = 1) \right|.$$

The *maximum PRF-advantage of* F is defined as

$$\mathbf{Adv}^{\mathrm{prf}}_F(q, \ell, \sigma) = \max_{\mathcal{A}} \mathbf{Adv}^{\mathrm{prf}}_F(\mathcal{A}),$$

where the maximum is taken over all adversaries $\mathcal{A}$ making at most q queries, each of length at most ℓ, and the total length of all queries at most σ, i.e., $\sigma \leq \ell q$.

Hash-then-RP Construction: Let $H : \mathcal{K} \times \mathcal{D} \to \mathfrak{B}$ be a keyed hash and $\boldsymbol{\pi}$ be an n-bit random permutation. The composition $\boldsymbol{\pi} \circ H_K$ is called the Hash-then-RP construction, where $K \leftarrow\!\!\$\, \mathcal{K}$. When $\boldsymbol{\pi}$ is replaced with $\boldsymbol{\rho}$, the resulting composition is called the Hash-then-RF. These constructions have been studied in [CW79, Sho96]. Many PRF constructions can be viewed as instances of Hash-then-RP/RF. For example, EMAC [BKR94, BdBB+95], ECBC, FCBC [BR00], LightMAC [LPTY16] and protected counter sum [Ber99]. Proposition 1 gives the PRF advantage for Hash-then-RP construction.

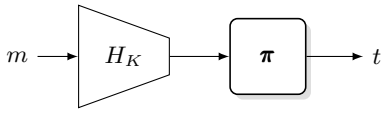

Figure 2.1: The Hash-then-RP paradigm.

Proposition 1. *Let H be a keyed block function with keyspace $\mathcal{K}$ and domain $\mathcal{D}$. Then, we have*

$$\mathbf{Adv}^{\mathrm{prf}}_{\boldsymbol{\pi}\circ H}(q,\ell,\sigma) \leq \mathsf{coll}_H(q,\ell,\sigma) + \frac{q(q-1)}{2^{n+1}}.$$

So, if H is an ϵ-universal hash function, then

$$\mathbf{Adv}^{\mathrm{prf}}_{\boldsymbol{\pi}\circ H}(q,\ell,\sigma) \leq \frac{q(q-1)}{2}\left(\epsilon + \frac{1}{2^n}\right).$$

We skip a formal proof here as Proposition 1 is a well-known result. The readers are referred to [GPR16] for a formal proof.

3 Revisiting Simplified PMAC

Description of sPMAC: Gaži et al. [GPR16] proposed a generalized version of PMAC, called sPMAC, to capture the underlying masking function for a wide class of PMAC variants. In what follows $\mathbb{N}$ denotes the set of all natural numbers.

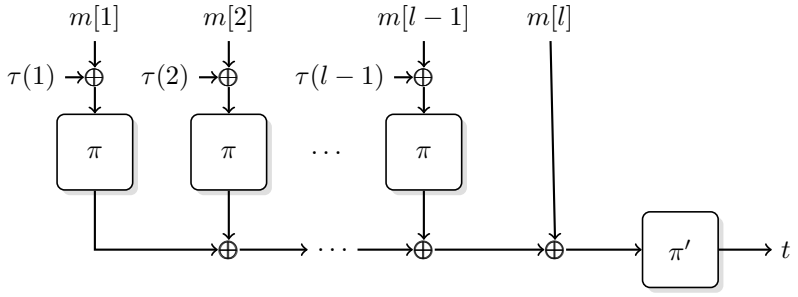

Figure 3.1: The simplified PMAC construction.

Definition 5 (sPHash)**.** For any permutation $\pi \in \mathsf{Perm}$ and a block-valued function $\tau \in \mathsf{Func}_{\mathbb{N}}$ (referred as masking function), we define the *simplified PMAC hash* or sPHash over the message space $\mathfrak{B}^+$ as follows:
for all $m := (m[1], \ldots, m[l]) \in \mathfrak{B}^l$,

$$\mathsf{sPHash}_{\pi,\tau}(m) := m[l] \oplus \bigoplus_{i=1}^{l-1} \pi(x_\tau(m,i)), \text{ where } x_\tau(m,a) := m[a] \oplus \tau(a). \tag{2}$$

Clearly, sPHash is just an identity function for a single block message.

Now, given two permutations $\pi, \pi' \in \mathsf{Perm}$ and a masking function $\tau \in \mathsf{Func}_{\mathbb{N}}$, the simplified PMAC or sPMAC construction (illustrated in Figure 3.1) is defined as follows:
for all $m \in \mathfrak{B}^+$,

$$\mathsf{sPMAC}_{\pi',\pi,\tau}(m) := \pi'\left(\mathsf{sPHash}_{\pi,\tau}(m)\right).$$

We call $K := (\pi', \pi, \tau)$ the key of sPMAC. A concrete variant of PMAC is determined whenever we fix a sampling mechanism of the key K.

sPMAC OVER ARBITRARY-LENGTH MESSAGES: For $m \in \{0,1\}^*$, we define

$$\overline{m} := m[1], \ldots, m[l] \xleftarrow{n} m$$

to be the function that partitions m into $l = \frac{|m\|10^i|}{n}$ blocks of size n bits, where i is the smallest non-negative integer such that $|m\|10^i|$ divisible by n. Note that, we make the required concatenation even if $|m|$ is divisible by n. sPMAC can be easily extended for any arbitrary-length message $m \in \{0,1\}^*$, as $\mathsf{sPMAC}(m) := \mathsf{sPMAC}(\overline{m})$. As the padding rule is injective, there is no loss of generality in ignoring the padding and assuming all message sizes are multiple of n.

PMAC VARIANTS FROM sPMAC: Now, we describe some variants of PMAC as instantiations of sPMAC by defining the sampling mechanism of the key $K = (\boldsymbol{\pi}, \boldsymbol{\pi}', \boldsymbol{\tau})$.

1. PMAC: We get the original PMAC [BR02] construction by setting $\boldsymbol{\pi} \leftarrow_{\$} \mathsf{Perm}$, $\boldsymbol{\pi}' = \boldsymbol{\pi}$, and $\boldsymbol{\tau}(i) = \gamma_i \cdot \boldsymbol{\pi}(0)$, where γ_i is the ith element of the Gray code sequence [Gra53, Rog04].

2. PMAC1: We get PMAC1 [Rog04] by setting $\boldsymbol{\pi} \leftarrow_{\$} \mathsf{Perm}$, $\boldsymbol{\pi}' = \boldsymbol{\pi}$, and $\boldsymbol{\tau}(i) = 2^i \cdot \boldsymbol{\pi}(0)$, where 2 is a fixed primitive element of the Galois field $\mathrm{GF}(2^n)$.

3. Gaži et al.'s variants: In [GPR16], Gaži et al. discussed two variants of PMAC. In both of the cases, $\boldsymbol{\pi}, \boldsymbol{\pi}' \leftarrow_{\$} \mathsf{Perm}$ and $\boldsymbol{\tau}$ is sampled independent of $\boldsymbol{\pi}, \boldsymbol{\pi}'$. The two choices of $\boldsymbol{\tau}$ are the following:
   (a) $\boldsymbol{\tau}$ is a uniform random function.
   (b) $\boldsymbol{\tau}$ is a 4-wise independent hash function.

4. Naito's variant of PMAC1: Naito proposed another variant of PMAC by setting $\boldsymbol{\pi}, \boldsymbol{\pi}' \leftarrow_{\$} \mathsf{Perm}$, and $\boldsymbol{\tau}(i) = 2^i \cdot L_1 \oplus 2^{3i} \cdot L_2$ where $L_1, L_2 \leftarrow_{\$} \mathfrak{B}$. In rest of the paper, we call this construction NPMAC.

N.B. In this paper, for the sake of simplicity, we pad all the messages (including the one whose length is a multiple of n). In the original PMAC(1), the last message block is padded only when it is incomplete (not a multiple of n) and is processed in a slightly different manner. However, our analyses are directly applicable to the actual PMAC(1) constructions.

UPPER BOUND ON THE PRF ADVANTAGE OF sPMAC: Any instance of sPMAC can be viewed as an instance of Hash-then-RP, as long as $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are sampled independently. Thus, the result of Hash-then-RP is not applicable for PMAC and PMAC1 as $\boldsymbol{\pi}' = \boldsymbol{\pi}$.

In this paper, we consider only those instances of sPMAC that follow the Hash-then-RP paradigm where $\boldsymbol{\pi}, \boldsymbol{\pi}', \boldsymbol{\tau}$ are all sampled independently. Moreover, $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ are random permutations and hence any PMAC variant (and its underlying hash) are completely determined once we fix a distribution for the masking function $\boldsymbol{\tau}$, say τ. We write sPHash_τ to represent $\mathsf{sPHash}_{\boldsymbol{\pi},\tau}$ and we write $\mathsf{sPMAC}_\tau(m) := \boldsymbol{\pi}'(\mathsf{sPHash}_\tau(m))$. We can restate Proposition 1 in context of PMAC variants as follows.

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(q,\ell,\sigma) \leq \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q,\ell,\sigma) + \frac{q(q-1)}{2^{n+1}} \tag{3}$$

$$\leq \frac{q(q-1)}{2} \cdot \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(\ell) + \frac{q(q-1)}{2^{n+1}} \tag{4}$$

Lower Bound on the PRF Advantage of sPMAC: Fix q distinct messages $m_1, \ldots, m_q$ such that

$$\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma) = \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m_1, \ldots, m_q).$$

In other words, the message tuple maximizes the collision probability. Now, we define a (non-adaptive) PRF distinguisher $\mathcal{A}$ for sPMAC that exploits collisions in sPMAC outputs.

1. $\mathcal{A}$ makes $2q$ queries, namely $m_1, m_1\|0^n, \ldots, m_q, m_q\|0^n$ to its oracle $\mathcal{O}$ (which is either $\mathsf{sPMAC}_{\boldsymbol{\tau}}$, i.e. the real oracle, or a random function, $\boldsymbol{\rho}$, i.e. the ideal oracle).
2. $\mathcal{A}$ returns 1, if for some $i \neq j$, $\mathcal{O}(m_i) = \mathcal{O}(m_j)$ as well as $\mathcal{O}(m_i\|0^n) = \mathcal{O}(m_j\|0^n)$, and 0 otherwise.

Note that, in case of real oracle, collision for m_i and m_j implies collision for $m_i\|0^n$ and $m_j\|0^n$ too. So, $\mathsf{Pr}(\mathcal{A}^{\mathsf{sPMAC}_{\boldsymbol{\tau}}} = 1) = \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m_1, \ldots, m_q)$, whereas, $\mathsf{Pr}(\mathcal{A}^{\boldsymbol{\rho}} = 1) \leq \frac{q(q-1)}{2^{2n+1}}$. So,

$$\begin{aligned}\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(\mathcal{A}) &\geq \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m_1, \ldots, m_q) - \frac{q(q-1)}{2^{2n+1}}. \\ &\geq \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma) - \frac{q(q-1)}{2^{2n+1}}. \end{aligned} \tag{5}$$

It is clear from Eq. (3) and (5) that $\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma)$ is a very close estimate for $\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}$, i.e., we have

$$\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma) - \frac{q(q-1)}{2^{2n+1}} \leq \mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(\mathcal{A}) \leq \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma) + \frac{q(q-1)}{2^{n+1}}. \tag{6}$$

In other words, $\left|\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(\mathcal{A}) - \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma)\right| \leq \frac{q(q-1)}{2^{n+1}}$.

3.1 Collision Analysis of sPMAC [GPR16]

We fix two distinct messages $m := (m[1], \ldots, m[l]), m' := (m'[1], \ldots, m'[l'])$ with number of blocks $l := l_m$ and $l' := l_{m'}$ respectively. We also assume $l \leq l'$. Let $m_{\mathrm{chop}} := (m[1], \ldots, m[l-1])$ denote the message m after removing the last block. Similarly, we write m'_{chop} for the message m'. Let

$$\mathcal{V} := \{(M, a) \mid M \in \{m, m'\}; 1 \leq a \leq l_M - 1\}$$

be called index set. For any masking function τ, recall the definition of x_τ (also referred as input function) from Eq. (2). x_τ can be viewed as a block function defined over $\mathcal{V}$. For a masking function τ, we write the multiset corresponding to all inputs for the chopped message m_{chop} as

$$\mathcal{X}_\tau(m_{\mathrm{chop}}) := \{x_\tau(m, 1), x_\tau(m, 2), \ldots, x_\tau(m, l-1)\}.$$

We similarly define $\mathcal{X}_\tau(m'_{\mathrm{chop}})$ for the message m' and $\mathcal{X}_\tau(m_{\mathrm{chop}}, m'_{\mathrm{chop}}) := \mathcal{X}_\tau(m_{\mathrm{chop}}) \cup \mathcal{X}_\tau(m'_{\mathrm{chop}})$. Note that $\mathcal{X}_\tau(m_{\mathrm{chop}}, m'_{\mathrm{chop}})$ actually depends on m_{chop} and m'_{chop}.

Definition 6 (cross-canceling masking function)**.** A masking function $\boldsymbol{\tau}$ is called cross-canceling with respect to m_{chop} and m'_{chop} if $\mathcal{X}_{\boldsymbol{\tau}}(m_{\mathrm{chop}}, m'_{\mathrm{chop}})$ is evenly repeated. Let

$$\theta_{\boldsymbol{\tau}}(m_{\mathrm{chop}}, m'_{\mathrm{chop}}) := \mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } (m_{\mathrm{chop}}, m'_{\mathrm{chop}})),$$

and $\theta_{\boldsymbol{\tau}}(\ell) := \max \theta_{\boldsymbol{\tau}}(m_{\mathrm{chop}}, m'_{\mathrm{chop}})$, where the maximum is taken over all distinct $m_{\mathrm{chop}}, m'_{\mathrm{chop}}$ with $l, l' < \ell$. $\theta_{\boldsymbol{\tau}}(\ell)$ is referred as the cross-cancellation probability of $\boldsymbol{\tau}$.

A proof of the following lemma is available in [GPR16, Lemma 2]. Similar result is also proved in [LPSY16, Proposition 1], albeit under a slightly different notational setup. We give another proof here for the sake of completeness.

Lemma 1 ([GPR16]). *For any random masking $\boldsymbol{\tau}$, we have*

$$\mathsf{coll}_{\mathit{sPHash}_{\boldsymbol{\tau}}}(\ell) \leq \theta_{\boldsymbol{\tau}}(\ell) + \frac{1}{2^n - 2\ell}.$$

Proof. Let m, m' be two distinct messages with $|m|, |m|' \leq \ell$. Now, the event $\mathsf{sPHash}_{\boldsymbol{\tau}}(m) = \mathsf{sPHash}_{\boldsymbol{\tau}}(m')$ can be divided in the following two disjoint events:

- $A : \mathsf{sPHash}_{\boldsymbol{\tau}}(m) = \mathsf{sPHash}_{\boldsymbol{\tau}}(m') \wedge \boldsymbol{\tau}$ is cross-canceling with respect to $(m_{\mathsf{chop}}, m'_{\mathsf{chop}})$
- $B : \mathsf{sPHash}_{\boldsymbol{\tau}}(m) = \mathsf{sPHash}_{\boldsymbol{\tau}}(m') \wedge \boldsymbol{\tau}$ is not cross-canceling with respect to $(m_{\mathsf{chop}}, m'_{\mathsf{chop}})$

The probability of event A can be bounded by $\theta_{\boldsymbol{\tau}}(m_{\mathsf{chop}}, m'_{\mathsf{chop}})$. Let us look at the event B. For simplicity of notation let us denote the multiset $\mathcal{X}_\tau(m_{\mathrm{chop}}, m'_{\mathrm{chop}})$ by $\mathcal{X}$. Then from Eq. (1) we have $\bigoplus_{x \in \mathcal{X}^o} \boldsymbol{\pi}(x) = m[l] \oplus m[l']$. Since $\mathcal{X}^o \neq \emptyset$, we can choose some $x_1 \in \mathcal{X}^o$ and bound $\mathsf{Pr}[B]$ as follows:

$$\mathsf{Pr}[B] \leq \mathsf{Pr}_{\boldsymbol{\pi}}\left[\boldsymbol{\pi}(x_1) = \oplus_{x \neq x_1} \boldsymbol{\pi}(x) \oplus m[l] \oplus m'[l']\right] \leq \frac{1}{2^n - l - l'} \leq \frac{1}{2^n - 2\ell} \tag{7}$$

In the first inequality we are considering x-values only from $\mathcal{X}^o$. The second inequality follows from probability of $\boldsymbol{\pi}(x_1)$ after we sample all other $\boldsymbol{\pi}$-values in a without replacement manner. Since we are left with exactly one choice among at least $2^n - l - l'$ many values here, we get the bound. The third inequality is obvious.

Therefore,

$$\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m, m') \leq \theta_{\boldsymbol{\tau}}(m, m') + \frac{1}{2^n - 2\ell}.$$

We get the required result by taking maximum over all m, m' such that $m \neq m'$ and $|m|, |m'| \leq \ell$ in both sides of the above inequality. □

Extension of Cross-Cancellation Probability over q Messages. In [GPR16], the idea of cross-cancellation is defined for two messages. Here, we extend the idea to more than two messages. For the sake of simplicity of notation we will write $\theta_{\boldsymbol{\tau}}(m, m')$ (and τ is cross-canceling with respect to m, m') instead of $\theta_{\boldsymbol{\tau}}(m_{\mathrm{chop}}, m'_{\mathrm{chop}})$ (and τ is cross-canceling with respect to $m_{\mathsf{chop}}, m'_{\mathsf{chop}}$). We say $\boldsymbol{\tau}$ to be cross-canceling with respect to m^q if $\boldsymbol{\tau}$ is cross-canceling with respect to m_i, m_j for some $1 \leq i < j \leq q$. Let

$$\theta_{\boldsymbol{\tau}}(m^q) := \mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } m^q),$$

and $\theta_{\boldsymbol{\tau}}(q, \ell, \sigma) := \max \theta_{\boldsymbol{\tau}}(m^q)$, where the maximum is taken over all q distinct messages each with at most $\ell - 1$ blocks, having at most $\sigma - q$ blocks altogether.

Lemma 2. *For any random masking $\boldsymbol{\tau}$, we have*

$$\theta_{\boldsymbol{\tau}}(q, \ell, \sigma) \leq \mathsf{coll}_{\mathit{sPHash}_{\boldsymbol{\tau}}}(q, \ell, \sigma) \leq \theta_{\boldsymbol{\tau}}(q, \ell, \sigma) + \frac{q(q-1)}{2(2^n - 2\ell)}.$$

Proof. Suppose, $m_1, \ldots, m_q$ are q messages for which $\theta_{\boldsymbol{\tau}}(m^q) = \theta_{\boldsymbol{\tau}}(q, \ell, \sigma)$. Let $\mathcal{T}$ denote the set of all realizable masking functions. Let $\mathcal{T}_{i,j} \subseteq \mathcal{T}$ denote the set of all cross-canceling masking functions with respect to (m_i, m_j). Then, $\theta_{\boldsymbol{\tau}}(m^q) := \mathsf{Pr}(\boldsymbol{\tau} \in \cup_{i<j} \mathcal{T}_{i,j})$. Let

$m'_i = m_i \| 0^n$ for $1 \le i \le q$. Now, for any $\tau \in \mathcal{T}_{i,j}$, $\mathsf{sPHash}_\tau(m'_i) = \mathsf{sPHash}_\tau(m'_j)$ holds (also denoted as $\mathsf{coll}_{i,j}$). So,

$$\theta_{\boldsymbol{\tau}}(q,\ell,\sigma) = \mathsf{Pr}(\boldsymbol{\tau} \in \cup_{i<j}\mathcal{T}_{i,j}) \le \mathsf{Pr}(\cup_{i<j}\mathsf{coll}_{i,j}) \le \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q,\ell,\sigma).$$

Now, we show the upper bound. We fix q distinct messages $m_1,\ldots,m_q$ such that $\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m^q) = \mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(q,\ell,\sigma)$. Let $\mu := \mathsf{Pr}(\boldsymbol{\tau}$ is cross-canceling with respect to $m^q)$.

$$\begin{aligned}
\mathsf{coll}_{\mathsf{sPHash}_{\boldsymbol{\tau}}}(m^q) &\le \mu + \sum_{\tau \in \mathcal{T}\setminus\cup_{i<j}\mathcal{T}_{i,j}} \mathsf{Pr}(\exists i<j, \boldsymbol{\pi}^{\oplus}(\mathcal{X}^o_\tau(m_i,m_j)) = m_i[l_i] \oplus m_j[l_j] \ \wedge\ \boldsymbol{\tau} = \tau) \\
&\le \mu + \sum_{\tau \in \mathcal{T}\setminus\cup_{i<j}\mathcal{T}_{i,j}} \mathsf{Pr}(\exists i<j, \boldsymbol{\pi}^{\oplus}(\mathcal{X}^o_\tau(m_i,m_j)) = m_i[l_i] \oplus m_j[l_j]) \times \mathsf{Pr}(\boldsymbol{\tau} = \tau) \\
&\le \mathsf{Pr}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } m^q) + \frac{q(q-1)}{2(2^n-2\ell)},
\end{aligned}$$

where the last inequality is obtained by conditioning on the output of $\boldsymbol{\pi}$ on all elements in $\mathcal{X}^o_\tau(m_i,m_j)$ except one. Note that this is possible only because $\mathcal{X}^o_\tau(m_i,m_j) \neq \emptyset$ since τ is not a cross-canceling function. □

Corollary 1. *For any random masking function $\boldsymbol{\tau}$, we have*

$$\begin{aligned}
\theta_{\boldsymbol{\tau}}(q,\ell,\sigma) - \frac{q^2}{2^{2n+1}} \le \mathbf{Adv}^{\mathrm{prf}}_{\mathit{sPMAC}_{\boldsymbol{\tau}}}(q,\ell,\sigma) &\le \theta_{\boldsymbol{\tau}}(q,\ell,\sigma) + \frac{q(q-1)}{2(2^n-2\ell)} + \frac{q(q-1)}{2^{n+1}} \\
&\le \frac{q(q-1)}{2}\cdot\theta_{\boldsymbol{\tau}}(\ell) + \frac{q(q-1)}{2(2^n-2\ell)} + \frac{q(q-1)}{2^{n+1}}.
\end{aligned}$$

Corollary 1 follows from Eq. (3) and Lemma 2 in combination with the observation that $\theta_{\boldsymbol{\tau}}(q,\ell,\sigma) \le \binom{q}{2}\theta_{\boldsymbol{\tau}}(\ell)$.

To achieve $O(q^2/2^n)$ bound, it is sufficient to show $\theta_{\boldsymbol{\tau}}(\ell) \le c/2^n$ for some constant c (should be independent of ℓ). Sometimes, it is possible to show this for a range of values of ℓ instead of all values of ℓ. Sometimes, it might be difficult to obtain ℓ-free bound for $\theta_{\boldsymbol{\tau}}(\ell)$. However, it might be possible to show ℓ-free bound for the $\theta_{\boldsymbol{\tau}}(q,\ell,\sigma)$ by considering all q messages together. In this case, first part of the above corollary could be used to obtain an ℓ-free security bound. When $\ell \le 2^{n-2}$, Corollary 1 is simplified to

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}} \le \frac{q^2}{2}\cdot\left(\theta_{\boldsymbol{\tau}}(\ell) + \frac{3}{2^n}\right). \tag{8}$$

Some Examples of Cross-Cancellation Probability: We list some known results on the cross-cancellation probability of some masking functions.

1. In [GPR16], Gaži et al. show the following bounds on cross-cancellation probability:
   (a) If $\boldsymbol{\tau}$ is a uniform random function, then $\theta_{\boldsymbol{\tau}}(\ell) \le 2^{1-n}$.
   (b) If $\boldsymbol{\tau}$ is a 4-wise independent hash function, then $\theta_{\boldsymbol{\tau}}(\ell) \le 2^{2-n}$.
2. For the masking function $\boldsymbol{\tau}(i) = 2^i \cdot L_i \oplus 2^{3i}\cdot L_2$, Naito proved the following result [Nai19, Section 4.2: Bounding p^2_{coll}] whenever $L_1, L_2 \leftarrow_{\$} \mathfrak{B}$:

$$\theta_{\boldsymbol{\tau}}(\ell) \le 2^{2-n}, \text{ while } \ell \le 2^{n/2}. \tag{9}$$

4 An Observation on Naito's PMAC Variant

In this section, we revisit a claim of [Nai19] regarding the cross cancellation probability of two powering-up maskings.

4.1 A Flaw and Its Effect on the Proof of NPMAC [Nai19]

As mentioned in section 3, Naito proved Eq. (9) with respect to the cross cancellation probability of two powering-up maskings. The proof relies on five cases [Nai19, Section 4.2: Type-1 to Type-5]. The most crucial and general of these cases is Type-5. Naito made the following claim with respect to this case.

Claim in [Nai19, Type-5 case in Section 4.2]: *The following system of equations, denoted* $(\mathcal{E})$*, in* L_1 *and* L_2 *such that* $\{i_1, i_2\} \neq \{i_3, i_4\}$*,*

$$(2^{i_1} \oplus 2^{i_2})L_1 \oplus (2^{3i_1} \oplus 2^{3i_2})L_2 = c_1$$
$$(2^{i_3} \oplus 2^{i_4})L_1 \oplus (2^{3i_3} \oplus 2^{3i_4})L_2 = c_2$$

has rank two (i.e. the equations are always linearly independent).

The author argues as follows: If the equations are not linearly independent then $2^{i_1} \oplus 2^{i_2} = 2^{i_3} \oplus 2^{i_4}$ and $2^{3i_1} \oplus 2^{3i_2} = 2^{3i_3} \oplus 2^{3i_4}$. From this, by using simple calculation, one can obtain $i_1 = i_2 = i_3 = i_4$. This leads to a contradiction of the assumption that $\{i_1, i_2\} \neq \{i_3, i_4\}$, and hence the linear dependence assumption is false. The author thus concludes that the system $(\mathcal{E})$ will always have rank 2. In other words, for fixed i_1, i_2, i_3, i_4, the system has a unique solution for the pair (L_1, L_2).

Flaw in the Argument: Unfortunately, linear dependency and consistency of the two equations over $\mathrm{GF}(2^n)$ can be equivalently written as

$$2^{i_1} \oplus 2^{i_2} = c \cdot (2^{i_3} \oplus 2^{i_4}) \tag{10}$$
$$2^{3i_1} \oplus 2^{3i_2} = c \cdot (2^{3i_3} \oplus 2^{3i_4}) \tag{11}$$

where $c_2 = c \cdot c_1$. Clearly, whenever $c \neq 1$, the claim on $(\mathcal{E})$ is not correct. In [Nai19], the author only considers the $c = 1$ case. Next, we show a concrete counterexample for this.

Counterexample for the Rank Claim: First, we can rewrite Eq. (10) and (11) as

$$(2^{i_1} \oplus 2^{i_2}) \cdot (2^{3i_3} \oplus 2^{3i_4}) = (2^{i_3} \oplus 2^{i_4}) \cdot (2^{3i_1} \oplus 2^{3i_2}) \tag{12}$$

We show a counterexample for $n = 16$. Similar examples can be constructed for other values of n as well. Consider the field $\mathrm{GF}(2^{16})$ generated by $x = 2$ with multiplication defined by the minimal polynomial $x^{16} + x^5 + x^3 + x + 1$. Using simple algebra one can show that $i_1 = 1$, $i_2 = 24$, $i_3 = 14$ and $i_4 = 18$ satisfies Eq. (12). Plugging in the same values in Eq. (11), one can get

$$c = 2^{12} \oplus 2^9 \oplus 2^8 \oplus 2^7 \oplus 2^6 \oplus 2^5 \oplus 2^2 \oplus 2 \oplus 1.$$

This proves that the system $(\mathcal{E})$ can be of rank 1 as well. And, the number of such i_1, i_2, i_3, i_4 is at least 1. Whereas, Naito incorrectly argues that the number of such quadruples is 0.

Effect on the Current Proof: The system $(\mathcal{E})$ is fixed once we fix the quadruple (i_1, i_2, i_3, i_4). In [Nai19], the number of i_1, i_2, i_3, i_4 indices corresponding to the system $(\mathcal{E})$ is bounded by $O(\ell^2)$ which can be further bounded by $O(2^n)$ (since $\ell \leq 2^{n/2}$). This bound is fine as long as the rank of system $(\mathcal{E})$ is 2, as this will mean that we get an overall cross-cancellation probability bound of $O(2^{-n})$. However, given the evidence that $(\mathcal{E})$ can have rank 1, a bound of $O(\ell^2)$ is not desirable, as it will result in an overall cross-cancellation probability bound of $O(\ell^2/2^n)$ which is worse than $O(\ell/2^n)$ bound for the existing PMAC.

4.2 Further Discussion on the Security of NPMAC

From previous discussions, it is clear that the question of ℓ-free security for NPMAC is far from resolved. Going by the existing proof strategy [Nai19], we get $\theta_{\boldsymbol{\tau}}(\ell) = O(\ell^2/2^n)$ bound. Looking ahead momentarily, Proposition 2 shows that we can achieve $O(\ell/2^n)$ for any $O(2^{-n})$-AXU masking function. This result also applies to NPMAC as the two powering-up maskings is obviously a $O(2^{-n})$-AXU. But, this is as far as we could reach. In what follows, we discuss some bottlenecks in resolving this question one way or another.

Let us denote the number of quadruples satisfying Eq. (12) by N. Our counterexample in the previous subsection shows that $N = \Omega(1)$ and due to Proposition 2 we can give a trivial upper bound of $N = O(\ell)$. Now, to prove or disprove the ℓ-free security claim we need an exact estimate of N.

We could neither construct a counterexample where $N = \Omega(\ell)$, nor show that $N = O(1)$. This indeed looks like a hard problem requiring an involved analysis of the properties of $\mathrm{GF}(2^n)$. Interestingly, a similar hardness remains for PMAC1 as well [LPSY16, GPR16] that involves a study of the additive subgroups (and their cosets) of $\mathrm{GF}(2^n)$.

Note that, $(\mathcal{E})$ is a simplified version of the actual system of equation that we have to analyze. In the actual system, c_1 and c_2 are not arbitrary. In fact, for some $M_1, M_2, M_3, M_4 \in \{m, m'\}$,

$$c_1 = M_1[i_1] \oplus M_2[i_2],\ c_2 = M_3[i_3] \oplus M_4[i_4], \text{ and thus, } c = \frac{M_3[i_3] \oplus M_4[i_4]}{M_1[i_1] \oplus M_2[i_2]}.$$

Clearly the simplification, though sufficient to discuss the flaw, could possibly degrade the bound as we count some inconsistent systems of equations as well. We say that a quadruple (i_1, i_2, i_3, i_4) is *valid* if the resulting system of equation $(\mathcal{E})$ is consistent. At the moment, we do not see any approach to exploit the exact nature of c to get a better estimate for the number of valid quadruples satisfying Eq. (12).
In summary, to prove or disprove the ℓ-free security of NPMAC, we have to bound:
The number, N, of valid quadruples (i_1, i_2, i_3, i_4) that satisfy

$$(2^{i_1} \oplus 2^{i_2}) \cdot (2^{3i_3} \oplus 2^{3i_4}) = (2^{i_3} \oplus 2^{i_4}) \cdot (2^{3i_1} \oplus 2^{3i_2}).$$

We leave it as an open problem to find an exact estimate for N, which in turn gives tight security bound for NPMAC. In fact, even a sub-optimal bound better than $\Omega(1)$ (in case of lower bound) or $O(\ell)$ (in case of upper bound), say in the order of a slowly growing function of ℓ, could be a great improvement.

5 Relaxing the Security Precondition for sPMAC

Gaži et al. [GPR16] showed that a 4-wise independent masking function is sufficient to achieve ℓ-free security bound up to $\ell \leq 2^{n/2}$. In this section, we relax the 4-wise independence condition to a weaker notion. Our relaxed notion of universality is inspired by the flaw discovered in section 4.

5.1 2-wise Almost XOR Universal Hash Function

We extend the definition of AXU hash functions to jointly consider two pairs of messages and their hash output differences.

Definition 7 (2-wise AXU)**.** A hash function H is called ϵ 2-wise AXU (or ϵ-2AXU) if for any distinct $\{m_1, m_2\}, \{m_3, m_4\}$ and $\delta_1, \delta_2 \in \mathfrak{B}$, we have

$$\mathsf{Pr}(H(K, m_1) \oplus H(K, m_2) = \delta_1 : K \leftarrow\!\!\$\, \mathcal{K}) \leq \epsilon,$$
$$\mathsf{Pr}(H(K, m_1) \oplus H(K, m_2) = \delta_1, H(K, m_3) \oplus H(K, m_4) = \delta_2 : K \leftarrow\!\!\$\, \mathcal{K}) \leq \epsilon^2.$$

Clearly, any ϵ-2AXU hash function is also an ϵ-AXU hash function. We usually expect $\epsilon = O(1/2^n)$ and hence the joint probability for the two linear equations is $O(1/2^{2n})$.

Mennink defined a very close variant, called AXU_3, in [Men18]. In that definition $m_3 = m_1$ (and hence $m_2 \neq m_4$). He also gave an example of AXU_3 (and its higher order variants) using finite field arithmetic.

2AXU is Strictly Weaker than 4-wise Independence: It is easy to see that a 4-wise independent hash function is 2^{-n}-2AXU. However, every 2AXU hash function need not be 4-wise independent. Consider the following example due to Naito [Nai, Nai20]. Similar example can also be found in [Men18] as an example of AXU_4 (see [Men18] for definition) hash function.

Example 2. Let $L_1, L_2, L_3 \xleftarrow{\$} \mathfrak{B}$. For a fixed primitive element 2 of $\text{GF}(2^n)$ and any i, let us define

$$\boldsymbol{\tau}(i) := 2^i \cdot L_1 \oplus 2^{2i} \cdot L_2 \oplus 2^{3i} \cdot L_3.$$

It can be easily shown that $\boldsymbol{\tau}$ is $O(2^{-n})$-2AXU. However, for any distinct i_1, i_2, i_3, i_4 and y_1, y_2, y_3, y_4 we cannot get probability $1/2^{4n}$ for the following event:

$$2^{i_j} \cdot L_1 \oplus 2^{2i_j} \cdot L_2 \oplus 2^{3i_j} \cdot L_3 = y_j, \forall j \in \{1, 2, 3, 4\}.$$

In other words, the above masking function is not 4-wise independent.

Remark 1. The two powering-up maskings used in [Nai19] is not 2^{-n}-2AXU hash. However, Naito addressed this issue in [Nai20] and proposed an alternate "three powering-up maskings" which is same as our example 2. He has given a dedicated proof for this construction whereas our proof for this one follows from our general treatment of 2AXU hash functions.

5.2 PRF Security of sPMAC

From Corollary 1, we know that the PRF advantage of sPMAC is bounded by the cross-cancellation probability of the underlying masking function. We have closely revisited all the existing proof strategies for upper bounding the cross-cancellation probability and have found a unified way to present all these proofs. This approach also helps in understanding the requirements from the masking function for achieving length-independent PRF advantage. We state two results unifying the proofs of existing and some new constructions. The proofs of these results is postponed to section 6.

Proposition 2. *Suppose $\boldsymbol{\tau}$ is ϵ-AXU. Then, $\theta_{\boldsymbol{\tau}}(\ell) \leq 2\ell\epsilon$. Hence, by using Corollary 1, we have*

$$\mathbf{Adv}^{\text{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(q, \ell, \sigma) \leq q^2\ell\epsilon + \frac{q^2}{2(2^n - 2\ell)} + \frac{q^2}{2^{n+1}}.$$

Proposition 2 gives the security bound for PMAC and PMAC1 when the outer permutation is replaced by an independent random permutation and the masking key is sampled independently. A dedicated analysis is required when we consider outer permutation same as the inner one and the masking key is derived from the permutation, like the original PMAC and PMAC1.

The bound in Proposition 2 is not ℓ-free as it has $q^2\ell\epsilon$ term (which came due to cross-cancellation probability). In the following result, we show how we can improve this term if we apply a stronger masking function. Gaži et al. [GPR16] proved a similar result for 4-wise independent masking function. However, we can easily extend their result to the weaker notion of 2AXU masking function.

Theorem 1. *Suppose $\boldsymbol{\tau}$ is ϵ-2AXU. Then, $\theta_{\boldsymbol{\tau}}(\ell) \leq \max\{2\epsilon, 4\ell^2\epsilon^2\}$. Hence, by using Corollary 1, we have*

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(q,\ell,\sigma) \leq \max\{q^2\epsilon, 2q^2\ell^2\epsilon^2\} + \frac{q^2}{2(2^n-2\ell)} + \frac{q^2}{2^{n+1}}.$$

So, when $\epsilon = 1/2^n$ and $\ell \leq 2^{\frac{n-1}{2}}$ then

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{sPMAC}_{\boldsymbol{\tau}}}(q,\ell,\sigma) \leq \frac{5q^2}{2^{n+1}}.$$

Theorem 1 also works (up to $\ell \leq 2^{n/2}$) for a uniform random masking function and 4-wise independent masking function as these are also $1/2^n$-2AXU hash functions. However, in case of uniform random function, a more precise analysis (as shown in [GPR16]) gives $\theta_{\boldsymbol{\rho}}(\ell) \leq 2/2^n$ for all values of ℓ.

Remark 2. Our result is a bit stronger than the result proved in [GPR16] as every 2AXU hash function need not be 4-wise independent hash function.

Remark 3. Theorem 1 gives an alternate proof of ℓ-free security for Naito's updated variant [Nai20] with three powering up masking (see example 2).

6 Proof of Theorem 1

Before we delve into the proofs of Proposition 2 and Theorem 1, we describe a graph-based description of input collisions that would help us to have some visual presentation of cross-canceling masking function.

6.1 Input Collision Graph and Covering Bound Lemma

Graph Notations: For a set V, let $[V]^2$ denote the set of all doubleton subsets of V. So, size of the set $[V]^2$ is $\binom{|V|}{2} := |V|(|V|-1)/2$. A graph G is a pair (V,E) where $E \subseteq [V]^2$. We call V and E the vertex and edge set of the graph, respectively. We also denote $V(G)$ and $E(G)$ to denote the vertex set and edge set of the graph G, respectively. An edge is an element $\{u,v\} \in E$ and we also say that u is adjacent to v. Given a graph $G = (V,E)$ and a subset $V' \subseteq V$, the subgraph restricted at V', denoted as $G(V')$, has vertex set V' and the edge set $[V']^2 \cap E$. A path from u to v of length t is a sequence of distinct elements $(w_0 := u, w_1, \ldots, w_t := v)$ such that w_{i-1} is adjacent to w_i for all $i \in [t]$. A component C (or connected component) is a subset of V such that for every $u,v \in C$ either $u = v$ or there is a path from u to v. A component C of a graph G is called clique if all pairs of the components are adjacent. We call a graph G *evenly partitioned* if all components of G have even sizes.

Input Collision Graph: Recall the index set $\mathcal{V} := \{(M,a) \mid M \in \{m,m'\}; 1 \leq a \leq l_M - 1\}$ for two distinct messages m and m' of length $l = l_m$ and $l' = l_{m'}$, respectively, such that $l \leq l'$. To each masking function τ, we associate a collision graph G_τ with the vertex set $\mathcal{V}$ such that any two vertices (M_1,a_1) and (M_2,a_2) are said to be *adjacent* if $x_\tau(M_1,a_1) = x_\tau(M_2,a_2)$. So an input collision graph is always disjoint union of cliques.

A graph G' over $\mathcal{V}$ is called τ-realizable if there is a realizable masking function τ such that $G_\tau = G'$. Let $\mathcal{G}$ be the set of all such realizable graphs. Among all realizable graphs, we are interested in some special graphs, namely evenly partitioned graph. Let $\mathcal{G}_{\mathrm{even}}$ be the set of all realizable graphs which are evenly partitioned. The following observation is straightforward from the definition of cross-canceling masking function.

Claim 1. *A masking function $\boldsymbol{\tau}$ is cross-canceling if and only if the induced input collision graph $G_{\boldsymbol{\tau}}$ is evenly partitioned.*

Due to Corollary 1, it is now sufficient to bound the probability to realize any evenly partitioned graph (equivalent to realizing a cross-canceling masking function). Now, we identify a subset of vertices for which restricted subgraph over that subset is evenly partitioned whenever the graph is evenly partitioned. Let

$$\mathcal{V}^{=} := \{(M, a) : M \in \{m, m'\},\ a \leq l, l',\ m[a] = m'[a]\}.$$

So, $(m, a) \in \mathcal{V}^{=}$ if and only if $(m', a) \in \mathcal{V}^{=}$. For any such (m, a), we obviously have $x_\tau(m, a) = x_\tau(m', a)$ for all masking functions τ (not necessarily cross-canceling masking function). Hence, for any realizable input collision graph G_τ, $\{(m, a), (m', a)\}$ is an edge of the graph and we call those edges *vertical* (all other edges will be non-vertical). On the other hand, if $(m, a) \notin \mathcal{V}^{=}$ then (m, a) and (m', a) are not adjacent whenever these are defined. Let $\mathcal{V}^{\neq} := \mathcal{V} \setminus \mathcal{V}^{=}$ and

$$I^{\neq} := \{a :\ \text{either } (m, a) \in \mathcal{V}^{\neq} \text{ or } (m', a) \in \mathcal{V}^{\neq}\}.$$

We can rewrite the set $I^{\neq}$ as union of the interval $[l + 1, l']$ (can be the empty set) and $\{a : a \leq l, l' \text{ and } m[a] \neq m'[a]\}$. As $m \neq m'$, we have $\mathcal{V}^{\neq} \neq \emptyset$. Given any graph G we denote $G^{\neq} := G(\mathcal{V}^{\neq})$, the subgraph restricted on the set of vertices $\mathcal{V}^{\neq}$.

Now any connected component of G_τ consists of a connected component of $G_\tau^{\neq}$ with some additional pairs of vertices from $\mathcal{V}^{=}$. Hence, we have the following result.

Claim 2. *For all masking functions $\boldsymbol{\tau}$, $G_{\boldsymbol{\tau}}$ is evenly partitioned if and only if $G_{\boldsymbol{\tau}}^{\neq}$ is evenly partitioned.*

Now, we explain a method by which we can obtain an upper bound on the cross-canceling probability $\theta_{\boldsymbol{\tau}}(\ell)$ or $\theta_{\boldsymbol{\tau}}(q, \ell, \sigma)$. Let $\mathcal{G}^{\neq}_{\text{even}}$ be the collection of all evenly partitioned realizable graphs over the vertex set $\mathcal{V}^{\neq}$. Due to above claim, this is same as the collection of all restricted subgraphs with vertex set $\mathcal{V}^{\neq}$ of all evenly partitioned realizable graphs.

Definition 8 (covering collection of edges)**.** Let $\mathcal{I}$ be some index set such that for every $i \in \mathcal{I}$ we have an edge set $E_i \subseteq [\mathcal{V}^{\neq}]^2$. The collection $\mathcal{E} := \{E_i : i \in \mathcal{I}\}$ is said to cover evenly partitioned graphs if for all $G \in \mathcal{G}^{\neq}_{\text{even}}$, there exists $i := i_G \in \mathcal{I}$ such that $E_i \subseteq E(G)$.

For any edge $e := \{(M_1, a_1), (M_2, a_2)\} \in [\mathcal{V}]^2$, we say that event $e(\boldsymbol{\tau})$ holds if

$$\boldsymbol{\tau}(a_1) \oplus \boldsymbol{\tau}(a_2) = c_e := M_1[a_1] \oplus M_2[a_2].$$

We extend the above definition to an edge set E as follows: An event $E(\boldsymbol{\tau})$ holds if for all edges $e \in E$, $e(\boldsymbol{\tau})$ holds. All these events are defined based on the randomness of $\boldsymbol{\tau}$ only and we simply write $\mathsf{Pr}(e)$ or $\mathsf{Pr}(E)$ to denote the probability that the corresponding event holds under the randomness of $\boldsymbol{\tau}$.

Lemma 3 (Covering Bound Lemma)**.** *Suppose $\{E_i : i \in \mathcal{I}\}$ covers evenly partitioned graphs, then we have*

$$\mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \textit{ is cross-canceling with respect to } (m, m')) \leq \sum_{i \in \mathcal{I}} \mathsf{Pr}(E_i)$$

Proof. Let $\mathcal{T}^*$ denote the set of all cross-canceling masking functions with respect to (m, m'). For every E_i, let $\mathcal{T}_i$ denote the set of all masking function τ such that $E_i \subseteq E(G_\tau^{\neq})$. Now, we claim that $\mathcal{T}^* \subseteq \cup_i \mathcal{T}_i$. For any $\tau \in \mathcal{T}^*$, G_τ is an evenly partitioned graph and hence (using Claim 2) for some i, $E_i \subseteq E(G_\tau^{\neq}) \subseteq E(G_\tau)$. Thus, $\tau \in \mathcal{T}_i$. So the claim holds. The result follows from union bound. □

6.2 Proof of Proposition 2

Let i be the smallest element in $I^{\neq}$. We use shorthand notation $e_i(v)$ and $e'_i(v)$ to denote edges $\{(m,i),v\}$ and $\{(m',i),v\}$, respectively, whenever these are defined. Let $\mathcal{V}_i^{\neq} := \mathcal{V}^{\neq} \setminus \{(m,i),(m',i)\}$.

As (m',i) has an edge for any evenly partitioned graph $G \in \mathcal{G}^{\neq}_{\text{even}}$, there must exist (M,j) with $j > i$ and $M \in \{m,m'\}$ such that (m',i) is adjacent to (M,j). So, we define the following collection of edge sets of size one.

$$\mathcal{E}_i := \{E_v := e'_i(v) : v \in \mathcal{V}_i^{\neq}\}.$$

From the above discussion, it is clear that this covers all evenly partitioned graphs. Now, using the fact that $\boldsymbol{\tau}$ is ϵ-AXU, we have $\mathsf{Pr}(E_{(M,j)}) = \mathsf{Pr}(\boldsymbol{\tau}(i) \oplus \boldsymbol{\tau}(j) = m'[i] \oplus M[j]) \leq \epsilon$ (since $j \neq i$). So, using the covering bound lemma (Lemma 3) we have

$$\mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } (m,m')) \leq \sum_{v \in \mathcal{V}_i^{\neq}} \mathsf{Pr}(E_v) \leq (l+l')\epsilon.$$

As $l, l' \leq \ell$, we have $\theta(\ell) \leq 2\ell\epsilon$. This completes the proof. □

6.3 Resuming the Proof of Theorem 1

Here, we first assume that $|I^{\neq}| > 2$, and we denote the first, second and third smallest elements of $I^{\neq}$ as i_1, i_2 and i_3, respectively. For $1 \leq j \leq 3$, $\mathcal{V}_j^{\neq} := \mathcal{V}^{\neq} \setminus \{(m,i_j),(m',i_j)\}$, and we use shorthand notation $e_j(v)$ and $e'_j(v)$ to denote edges $\{(m,i_j),v\}$ and $\{(m',i_j),v\}$, respectively, whenever these are edges over $\mathcal{V}$ (they may not be edge as some of the vertices may not be present in $\mathcal{V}$).

In the previous proof for AXU masking function, edge sets are singleton and hence the probability for any such edge set can be at best $O(1/2^n)$ (as we deal with a single equation). Now, we are considering doubleton edge sets, hoping that probability to realize any edge set is about $O(1/2^{2n})$ (as we assume stronger masking function), to achieve better security. Consider the following collections of doubleton edge sets:

1. $\mathcal{E}_1 := \{\{e'_1(M,i_2), e'_3(v)\} : v \in \mathcal{V}_3^{\neq}, M \in \{m,m'\}\}$,
2. $\mathcal{E}_2 := \{\{e'_1(M_1,j_1), e'_2(M_2,j_2)\} : (M_1,j_1) \in \mathcal{V}_1^{\neq} \cap \mathcal{V}_2^{\neq}, (M_2,j_2) \in \mathcal{V}_2^{\neq}\}$.

We claim that the collection $\mathcal{E} := \mathcal{E}_1 \cup \mathcal{E}_2$ is a covering collection of edges. Fix any evenly partitioned graph G over $\mathcal{V}^{\neq}$. The vertex (m',i_1) should be adjacent to some other vertex.

Case 1: Suppose, (m',i_1) is adjacent to (M,i_2) then the vertex (m',i_3) should be adjacent to (M,j) for some $j \neq i_3$. So, we can use an appropriate edge set from $\mathcal{E}_1$.

Case 2: Suppose, (m',i_1) is adjacent to (M,j) for some $M \in \{m,m'\}$ and $j \geq i_3$. Then, (m',i_2) should be adjacent to (M,j) for some $j \neq i_2$. So, we can use an appropriate collection from $\mathcal{E}_2$.

Thus, $\mathcal{E}$ is indeed a covering collection of edges. Now, we fix any edge set $E := \{e'_1(M_1,i_2), e'_3(M_2,j)\} \in \mathcal{E}_1$ where $j \neq i_3$. Then, for $c_1 = m'[i_1] \oplus M_1[i_2]$ and $c_2 = m'[i_3] \oplus M_2[j]$, we have

$$\mathsf{Pr}(E) = \mathsf{Pr}(\boldsymbol{\tau}(i_1) \oplus \boldsymbol{\tau}(i_2) = c_1, \boldsymbol{\tau}(i_3) \oplus \boldsymbol{\tau}(j) = c_2) \leq \epsilon^2,$$

where the inequality follows from the definition of ϵ-2AXU. Similarly, for any edge set $E \in \mathcal{E}_2$, one can show that $\mathsf{Pr}(E) \leq \epsilon^2$. Note that $|\mathcal{E}_1| \leq 2(l+l')$ and $|\mathcal{E}_2| \leq (l+l'-2)\cdot(l+l'-4)$. So, $|\mathcal{E}| \leq (l+l')^2 \leq 4\ell^2$. By using the covering bound Lemma (Lemma 3), we have

$$\mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } (m,m')) \leq \sum_{E \in \mathcal{E}} \mathsf{Pr}(E) \leq 4\ell^2\epsilon^2.$$

Now, the only remaining case is $|I^{\neq}| = 2$ ($|I^{\neq}|$ cannot be 1 as this would contradict the existence of evenly partitioned graph). In this case, we have only two possibilities of evenly partitioned graphs, each occurring with at most ϵ probability (using ϵ-2AXU). So, we have

$$\mathsf{Pr}_{\boldsymbol{\tau}}(\boldsymbol{\tau} \text{ is cross-canceling with respect to } (m, m')) \leq 2\epsilon.$$

The result follows by combining the two cases for $|I^{\neq}|$. □

7 PMAC2: A Simple Variant of PMAC1

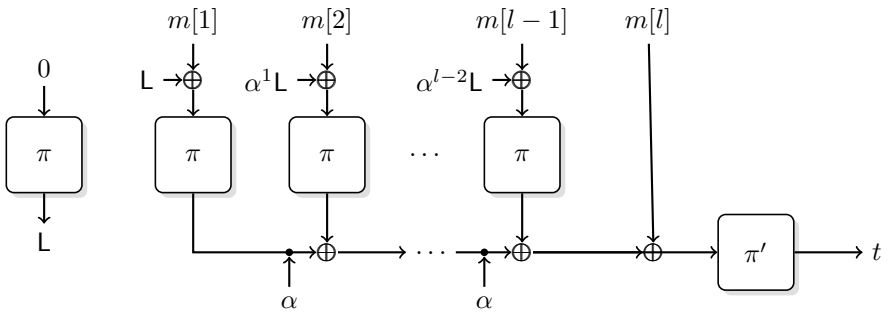

Figure 7.1: PMAC2: A message m is padded with 10^* to get $m[1]||m[2]||\dots m[l]$ where each $m[i]$ is an n-bit string. L is obtained as $\boldsymbol{\pi}(0)$ where $\boldsymbol{\pi} \leftarrow\!\!\$\ \mathsf{Perm}$. Here α is a primitive element of the field $GF(2^n)$.

Now we propose a simple variant of PMAC1 which we call PMAC2 (see Fig. 7.1). Given any message $m' \in \{0,1\}^*$ we append a bit 1 followed by a smallest sequence of zeros so that the padded message has size multiple of n. Let $m := (m[1], \dots, m[l]) \in \mathfrak{B}^l$ be a padded message. As it is an injective padding, we define the construction after the padding. Let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be two independent random permutations (for a real construction we use a block cipher instantiated by two independent keys). We compute the final output of $\mathsf{PMAC2}(m)$ as follows:

$\mathsf{PMAC2}_{\boldsymbol{\pi},\boldsymbol{\pi}'}(m)$

1 : **Input:** $m = m[1]\| \dots \|m[l]$
2 : $L \leftarrow \boldsymbol{\pi}(0)$
3 : **for** $i = 1$ *to* $l-1$, **do** $x[i] \leftarrow m[i] \oplus \alpha^{i-1} \cdot L$
4 : $H_{\boldsymbol{\pi}} \leftarrow m[l] \oplus \bigoplus_{i=1}^{l-1} \alpha^{l-i-1} \boldsymbol{\pi}(x[i])$
5 : **return** $\boldsymbol{\pi}'(H_{\boldsymbol{\pi}})$

Theorem 2. ***(Main Theorem: Bound for Hash Collision Probability of PMAC2)***

$$\mathsf{coll}_H(q, \ell, \sigma) = \frac{q^2 + \sigma}{2^n} + \mu$$

where $\mu \leq \begin{cases} \frac{q}{2^{n/2}} & \text{if } \ell \leq 2^{n/4} \\ \frac{\sigma^{1.5}}{2^n} & \text{if } 2^{n/4} < \ell \leq 2^{n-2}. \end{cases}$

We prove this theorem in the next subsection. The PRF-advantage of our construction will follow from hash-then-prp result:

$$\mathbf{Adv}^{\mathrm{prf}}_{\mathsf{PMAC2}}(q,\ell,\sigma) \le \mathsf{coll}_H(q,\ell,\sigma) + \frac{q^2}{2^{n+1}}.$$

Remark 4. The original proof for PMAC works perfectly in the case of PMAC2 and hence the security of PMAC2 is also bounded by the security bound of PMAC. Our result gives a different bound of PMAC2 which essentially gives tighter bounds in the cases of $\ell < 2^{n/4}$ and $\ell \ge 2^{n/4}$ such that $\ell < q$. In all other cases we can take the usual bound for PMAC. To be precise, we can always choose the minimum between the usual bound for PMAC and the bound obtained here.

7.1 Proof of Theorem 2

Let $m^q = (m_1, \ldots, m_q)$ be a q-tuple of distinct messages. Let $\ell_i = \|m_i\|$, $\ell := \max_i \ell_i$ and $\sigma := \sum_i \ell_i$. For simplicity we will write $H(m)$ instead of $H_{\boldsymbol{\pi}}(m)$ for any message m. We want to bound $\mathsf{coll}(m^q) := \mathsf{Pr}_{\boldsymbol{\pi} \xleftarrow{\$} \mathsf{Perm}}[\exists i \ne j,\ H(m_i) = H(m_j)]$. Note that we use the masking function $\tau_L(i) := \alpha^{i-1} \cdot L$ where $L = \boldsymbol{\pi}(0)$. For every $i \ne j$, we have already defined a graph $G_{\tau_L}(m_i, m_j)$ (defined previously as G_{τ_L} for any block function τ and a pair of distinct messages m, m'). Note that, we explicitly associated the graph G_{τ_L} with the corresponding message pair (m_i, m_j) as we are dealing with multiple message pairs. We will drop this parametrization whenever the message pair is known from the context. Here, for simplicity, we will denote any vertex by (k, a) instead of (m_k, a). G_{τ_L} is essentially a disjoint union of cliques. For any clique C in G_{τ_L} we define

$$\beta_C := \bigoplus_{(k,a)\in C} \alpha^{\ell_k - 1 - a}$$

Definition 9. A masking function τ_L (or simply L) is *cross linear canceling* for some $i \ne j$, if $\beta_C = 0$ for every clique C in $G_L(m_i, m_j)$. We define

$$\theta'(m^q) := \mathsf{Pr}_L[\exists i \ne j,\ \tau_L \text{ is cross linear cancelling for } i, j].$$

Avoiding Zero Input. We first avoid zero block as an input of $\boldsymbol{\pi}$ since it already appears to define our masking key L. We define the following event:

$$\mathsf{bad}_0 : \exists i, a,\ x_i[a] = 0$$

Clearly, $\mathsf{Pr}[\mathsf{bad}_0] \le \frac{\sigma}{2^n}$ as for every (i, a), $\mathsf{Pr}(x_i[a] = 0) = 1/2^n$.

It is easy to see that if L is cross linear canceling for i, j, then $H(m_i) = H(m_j)$. Therefore, a similar statement like Lemma 1 holds:

Lemma 4.

$$\mathsf{coll}(m^q) \le \theta'(m^q) + \frac{q^2}{2(2^n - 2\ell)} + \frac{\sigma}{2^n}$$

Proof. Let $H(m_i) = H(m_j)$ for some $i < j \in [q]$. Then one of the following three events must happen:

- bad_0
- $A(i, j) : \tau_L$ is cross linear canceling for $i, j \wedge H(m_i) = H(m_j)$
- $B(i, j) : \tau_L$ is not cross linear canceling for $i, j, \wedge H(m_i) = H(m_j) \ \wedge \ \neg\mathsf{bad}_0$

Therefore,

$$\begin{aligned}\mathsf{coll}(m^q) &\le \mathsf{Pr}[\cup_{i<j} A(i,j)] + \mathsf{Pr}[\cup_{i<j} B(i,j)] + \mathsf{Pr}[\mathsf{bad}_0] \\ &\le \theta'(m^q) + \mathsf{Pr}[\cup_{i<j} B(i,j)] + \frac{\sigma}{2^n}\end{aligned} \tag{13}$$

since $\mathsf{Pr}[\cup_{i<j} A(i,j)] \le \theta'(m^q)$ and $\mathsf{Pr}[\mathsf{bad}_0] \le \frac{\sigma}{2^n}$.

Let us now consider the event $B(i,j)$. That τ_L is not cross linear canceling for i,j implies that there exists a component C_1 in the graph $G_{\tau_L}(m_i,m_j)$ such that $\beta_{C_1} \neq 0$. For any component C of the graph, we get a unique value, say $x(C)$ such that $x_k(a) = x(C)$ for any $(k,a) \in C$. Note that for any two distinct components C and C', $x(C) \neq x(C')$. Thus

$$H(m_i) = H(m_j) \iff \beta_{C_1} \cdot \boldsymbol{\pi}(x(C_1)) = \bigoplus_{C \neq C_1} \beta_C \cdot \boldsymbol{\pi}(x(C)) \oplus m[\ell_i] \oplus m[\ell_j].$$

With the assumption $\neg\mathsf{bad}_0$, we can bound the probability of $B(i,j)$ using the randomness of $\boldsymbol{\pi}(x(C_1))$ (since $\beta_{C_1} \neq 0$) after we sample $\boldsymbol{\pi}$-values for all other components in a without replacement manner. Since the maximum number of components in $G_{\tau_L}(m_i,m_j)$ is $\ell_i + \ell_j$, we get

$$\mathsf{Pr}[B(i,j)] \le \frac{1}{2^n - \ell_i - \ell_j} \le \frac{1}{2^n - 2\ell} \tag{14}$$

Therefore, applying union bound on $\cup_{i<j} B(i,j)$ we get the required bound for $\mathsf{coll}(m^q)$ directly from Eq. (13). □

Now, it suffices to bound $\theta'(m^q)$. For the time being we assume that $\ell_i = \ell$ for all i. Later we will relax the assumption and complete our proof.

Lemma 5.

$$\theta'(m^q) \le \frac{\min\{q\ell^2, 2q^2\ell\}}{2^{n+1}}$$

Proof. For any $i < j$, we let $a_{i,j} := \min I^{\neq}(m_i,m_j)$. Consider the following two events:

- $\mathsf{bad}_1 : \exists i \in [q], \exists b,c \in [\ell-1], b < c,$ such that $x_i[b] = x_i[c]$
- $\mathsf{bad}_2 : \exists i < j \in [q], \exists b \in [\ell-1], b \neq a_{i,j},$ such that $(x_i[a_{i,j}] = x_j[b]) \vee (x_i[a_{i,j}] = x_i[b])$

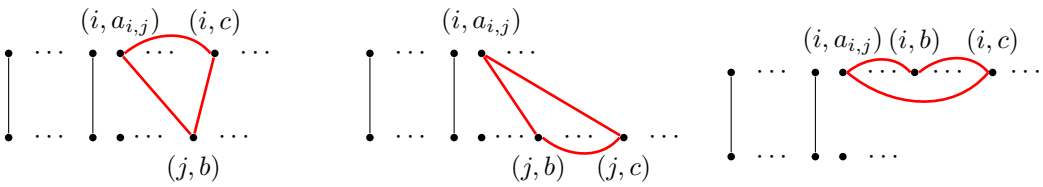

Figure 7.2: One of these is a necessary subgraph of a cross linear canceling graph for two messages with same block-lengths. A red or (solid) black line between two nodes signifies equality between them. Red is used when two blocks with different positions collide. Black is used when two blocks with same position collide.

Using randomness of L, we can easily bound the probability of the above two bad events.

$$\begin{aligned}\mathsf{Pr}[\mathsf{bad}_1] &\le \frac{q\ell^2}{2^{n+1}} \\ \mathsf{Pr}[\mathsf{bad}_2] &\le \frac{2q^2\ell}{2^{n+1}}\end{aligned} \tag{15}$$

We claim that if L is cross linear canceling for some message pair (m_i, m_j), then both bad events bad_1 and bad_2 hold. We first note that $2^n - 1 > \ell > 1$. Now consider the clique C of $G_{\tau_L}(m_i, m_j)$ that contains $(i, a_{i,j})$. From the definition of β_C and the assumption that $\ell_i = \ell_j$, we note that β_C can be zero only if C contains at least three vertices. Figure 7.2 illustrates all possible types of sub-clique of C, containing exactly three vertices, one of which is $(i, a_{i,j})$. It is obvious to see that at least two of the vertices must appear in the same query, whence we establish that bad_1 holds. Further, Figure 7.2 shows all possible way in which $x_i[a_{i,j}]$ is connected to some vertex, which establishes that bad_2 must hold. This validates our claim. The proof follows from Eq. (15). □

Handling Different Length Queries:

Claim 3. *If two messages m_i and m_j are of different length then τ is not cross linear canceling for i, j.*

To show this, suppose τ is cross linear canceling for i, j. Without any loss of generality assume $\|m_i\| > \|m_j\|$. Then each β_C must be 0. Thus the sum over all β_C where C is a clique in G_τ must also be 0. Note that

$$\bigoplus_C \beta_C = \bigoplus_{i=\|m_j\|}^{\|m_i\|-1} \alpha^i$$

which can never be 0 since α is a primitive element of $GF(2^n)$.

Now, we group together all the messages with same block-lengths. Precisely, for any $l \in [\ell]$, we define the following notations:

$$S_l := \{i \in [q] : \|m_i\| = l\}; \quad s_l := |S_l|;$$

Note that $\sum_l s_l = q, \sum_l s_l l = \sigma$.

Moreover, for any $l \in [\ell]$, we define $m^{S_l} := (m_{i_1}, \ldots, m_{i_{s_l}})$ where $\{i_1, \ldots, i_{s_l}\}$ denotes the set S_l in ascending order. Therefore,

$$\theta'(m^{S_l}) = \mathsf{Pr}[\exists i \neq j \in S_l \text{ s.t. } L \text{ is cross linear cancelling for } i, j].$$

Using Claim 3 and Lemma 5 we have

$$\theta'(m^q) \leq \sum_l \theta'(m^{S_l}) \leq \mu := \sum_l \mu_l \text{ where } \mu_l := \frac{\min\{s_l l^2, 2s_l^2 l\}}{2^{n+1}}. \tag{16}$$

In the remainder, we derive upper bounds on μ depending upon the range of ℓ values. First, consider $\ell \leq 2^{n/4}$. In this case, we have $\mu_l \leq \frac{s_l}{2^{n/2}}$ which implies

$$\mu \leq \frac{q}{2^{n/2}}. \tag{17}$$

Now, consider $\ell > 2^{n/4}$. Using the fact that for positive reals a and b, $\sqrt{ab} \geq \min\{a, b\}$, we have

$$\begin{aligned} \mu = \sum_l \mu_l &\leq \sum_l \frac{\sqrt{2}(s_l l)^{1.5}}{2^{n+1}} \\ &\leq \frac{\sigma^{1.5}}{2^n}, \end{aligned} \tag{18}$$

where the second inequality follows from the fact that $\Sigma_i a_i^r \leq (\Sigma_i a_i)^r$ for positive a_i and $r > 1$, and $\Sigma_l s_l l = \sigma$. Theorem 2 can be proved by plugging in the suitable values of μ from the above equations in Lemma 4, assuming $\ell \leq 2^{n-2}$.

8 Conclusion and Future Works

In this paper, we revisited some difficulties in designing a PMAC variant that has length-independent security bound $O(q^2/2^n)$ up to $\ell < 2^{n/2}$. Particularly, we took a closer look at a recent PMAC variant by Naito [Nai19] that claims to have length-independent security bound. We showed that the security proof of this construction has a non-trivial gap which is not easy to fix. Indeed, we pose it as an open problem to prove or disprove the ℓ-free security bound of $O(q^2/2^n)$ for Naito's construction. Apparently, this problem could be as hard as a similar problem posed in context of PMAC1 [Rog04]. On a positive note, we show that 2AXU (see section 5) masking function is sufficient to achieve length-independent security up to $\ell < 2^{n/2}$. This is a relaxation from the 4-wise independence condition used in [GPR16]. Finally, we proposed a simple variant of PMAC1, called PMAC2, that achieves ℓ-free security up to $\ell \leq 2^{n/4}$. For the range $2^{n/4} < \ell \leq 2^{n-2}$, PMAC2 still achieves ℓ-free security while $\sigma < 2^{2n/3}$.

Acknowledgements

We would like to thank Jooyoung Lee for his insightful comments on this paper. We would also like to thank the anonymous reviewers of ToSC Volume 2021 Issue 2 for their valuable comments and suggestions. We would like to thank Yusuke Naito for giving his valuable time to go through our findings and sharing his feedback. Ashwin Jha's work was carried out in the framework of the French-German-Center for Cybersecurity, a collaboration of CISPA and LORIA. Bishwajit Chakraborty, Soumya Chattopadhyay and Mridul Nandi are supported by the project "Study and Analysis of IoT Security" under Government of India at R.C.Bose Centre for Cryptology and Security, Indian Statistical Institute, Kolkata.

References

[BdBB+95] A. Berendschot, B. den Boer, J. Boly, A. Bosselaers, J. Brandt, D. Chaum, I. Damgård, M. Dichtl, W. Fumy, M. van der Ham, C. Jansen, P. Landrock, B. Preneel, G. Roelofsen, P. de Rooij, and J Vandewalle. Final Report of RACE Integrity Primitives. *Lecture Notes in Computer Science, Springer-Verlag, 1995*, 1007, 1995.

[Ber99] Daniel J. Bernstein. How to stretch random functions: The security of protected counter sums. *J. Cryptology*, 12(3):185–192, 1999.

[BGM04] Mihir Bellare, Oded Goldreich, and Anton Mityagin. The power of verification queries in message authentication and authenticated encryption. *IACR Cryptology ePrint Archive*, 2004:309, 2004.

[BKR94] Mihir Bellare, Joe Kilian, and Phillip Rogaway. The security of cipher block chaining. In *Advances in Cryptology – CRYPTO '94, Proceedings*, pages 341–358, 1994.

[BR00] John Black and Phillip Rogaway. CBC macs for arbitrary-length messages: The three-key constructions. In *Advances in Cryptology - CRYPTO '00, Proceedings*, pages 197–215, 2000.

[BR02] John Black and Phillip Rogaway. A Block-Cipher Mode of Operation for Parallelizable Message Authentication. In *Advances in Cryptology - EUROCRYPT 2002, Proceedings*, pages 384–397, 2002.

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

[DDNP18] Nilanjan Datta, Avijit Dutta, Mridul Nandi, and Goutam Paul. Double-block hash-then-sum: A paradigm for constructing bbb secure prf. *IACR Trans. Symmetric Cryptol.*, 2018(3):36–92, 2018.

[DJN17] Avijit Dutta, Ashwin Jha, and Mridul Nandi. A new look at counters: Don't run like marathon in a hundred meter race. *IEEE Trans. Computers*, 66(11):1851–1864, 2017.

[EMST76] William F. Ehrsam, Carl H. W. Meyer, John L. Smith, and Walter L. Tuchman. Message Verification and Transmission Error Detection by Block Chaining. Patent 4074066, USPTO, 1976.

[GGM84] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions (extended abstract). In *25th Annual Symposium on Foundations of Computer Science - FOCS '84, Proceedings*, pages 464–479, 1984.

[GPR16] Peter Gazi, Krzysztof Pietrzak, and Michal Rybár. The exact security of PMAC. *IACR Trans. Symmetric Cryptol.*, 2016(2):145–161, 2016.

[Gra53] Frank Gray. Pulse code communication. Patent 2632058, USPTO, 1953.

[JN16a] Ashwin Jha and Mridul Nandi. Revisiting structure graphs: Applications to CBC-MAC and EMAC. *J. Mathematical Cryptology*, 10(3-4):157–180, 2016.

[JN16b] Ashwin Jha and Mridul Nandi. Revisiting structure graphs: Applications to CBC-MAC and EMAC. *IACR Cryptology ePrint Archive*, 2016:161, 2016.

[KLL20] Seongkwang Kim, ByeongHak Lee, and Jooyoung Lee. Tight security bounds for double-block hash-then-sum macs. In *Advances in Cryptology - EUROCRYPT 2020, Proceedings, Part I*, pages 435–465, 2020.

[LPSY16] Atul Luykx, Bart Preneel, Alan Szepieniec, and Kan Yasuda. On the influence of message length in pmac's security bounds. In *Advances in Cryptology - EUROCRYPT 2016, Proceedings, Part I*, pages 596–621, 2016.

[LPTY16] Atul Luykx, Bart Preneel, Elmar Tischhauser, and Kan Yasuda. A MAC mode for lightweight block ciphers. In *Fast Software Encryption - 23rd International Conference, FSE 2016, Revised Selected Papers*, volume 9783, pages 43–59, 2016.

[Men18] Bart Mennink. Towards tight security of cascaded LRW2. In *Theory of Cryptography - TCC '18. Proceedings, Part II*, pages 192–222, 2018.

[MM07] Kazuhiko Minematsu and Toshiyasu Matsushima. New Bounds for PMAC, TMAC, and XCBC. In *Fast Software Encryption – FSE '07, Revised Selected Papers*, pages 434–451, 2007.

[Nai] Yusuke Naito. Personal communication.

[Nai17] Yusuke Naito. Blockcipher-based macs: Beyond the birthday bound without message length. In *Advances in Cryptology - ASIACRYPT 2017, Part III*, pages 446–470, 2017.

[Nai19] Yusuke Naito. The exact security of PMAC with two powering-up masks. *IACR Trans. Symmetric Cryptol.*, 2019(2):125–145, 2019.

[Nai20] Yusuke Naito. The exact security of PMAC with three powering-up masks. *IACR Cryptol. ePrint Arch.*, 2020:731, 2020.

[NIS01] NIST. Announcing the ADVANCED ENCRYPTION STANDARD (AES). FIPS 197, National Institute of Standards and Technology, U. S. Department of Commerce, 2001.

[NM08] Mridul Nandi and Avradip Mandal. Improved security analysis of PMAC. *J. Mathematical Cryptol.*, 2(2):149–162, 2008.

[Rog04] Phillip Rogaway. Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In *Advances in Cryptology - ASIACRYPT '04, Proceedings*, pages 16–31, 2004.

[Sho96] Victor Shoup. On fast and provably secure message authentication based on universal hashing. In *Advances in Cryptology - CRYPTO '96. Proceedings*, pages 313–328, 1996.

[Sho04] Victor Shoup. Sequences of games: a tool for taming complexity in security proofs. *IACR Cryptology ePrint Archive*, 2004:332, 2004.

[Yas11] Kan Yasuda. A New Variant of PMAC: Beyond the Birthday Bound. In *Advances in Cryptology - CRYPTO 2011, Proceedings*, pages 596–609, 2011.